

# Fluctuation geometry: A counterpart approach of inference geometry

**L. Velazquez**

Departamento de Física, Universidad Católica del Norte, Av. Angamos 0610, Antofagasta, Chile.

**Abstract.** Starting from an axiomatic perspective, *fluctuation geometry* is developed as a counterpart approach of inference geometry. This approach is inspired on the existence of a notable analogy between the general theorems of *inference theory* and the *general fluctuation theorems* associated with a parametric family of distribution functions  $dp(I|\theta) = \rho(I|\theta)dI$ , which describes the behavior of a set of *continuous stochastic variables* driven by a set of control parameters  $\theta$ . In this approach, statistical properties are rephrased as purely geometric notions derived from the *Riemannian structure* on the manifold  $\mathcal{M}_\theta$  of stochastic variables  $I$ . Consequently, this theory arises as an alternative framework for applying the powerful methods of differential geometry for the statistical analysis. Fluctuation geometry has direct implications on statistics and physics. This geometric approach inspires a Riemannian reformulation of Einstein fluctuation theory as well as a geometric redefinition of the information entropy for a continuous distribution.

PACS numbers: 02.50.-r; 02.40.Ky; 05.45.-a; 02.50.Tt

Keywords: Geometrical methods in statistics; Fluctuation theory; Information theory

## 1. Introduction

Inference theory supports the introduction of a Riemannian distance notion [1]:

$$ds_F^2 = g_{\alpha\beta}(\theta)d\theta^\alpha d\theta^\beta \quad (1)$$

to characterize the *statistical distance* between two close members of a generic parametric family of distribution functions:

$$dp(I|\theta) = \rho(I|\theta)dI. \quad (2)$$

Here, the metric tensor  $g_{\alpha\beta}(\theta)$  is provided by the so-called *Fisher's information matrix* [2]. The existence of this type of Riemannian formulation was pioneering suggested by Rao [3], which is hereafter referred to as *inference geometry*<sup>‡</sup>. Inference geometry provides very strong tools for proving results about statistical models, simply by considering them as well-defined geometrical objects. As expected, inference theory and its geometry have a direct application in those physical theories with a statistical apparatus as statistical mechanics and quantum theories. A curious example is the so-called *extreme physical information principle*, proposed by Frieden in 1998, which claims that all physical laws can be derived from purely inference arguments [4]. Inference geometry

<sup>‡</sup> This approach is referred to as *Riemannian geometry on statistical manifolds* in the literature. However, the denomination *inference geometry* is employed here to avoid the ambiguity with *fluctuation geometry*, which is also a Riemannian geometry on statistical manifold.

has been adapted to the mathematical apparatus of quantum mechanics [5, 6, 7, 8]. In fact, modern interpretations of uncertainty relations are inspired on their arguments [9, 10, 11]. Inference geometry has been successfully employed in statistical mechanics to study phase transitions [12, 13], as well as in the framework of so-called thermodynamics geometry [14].

The goal of this paper is to show that the parametric family of distribution functions (2) supports the introduction of an alternative Riemannian distance notion:

$$ds^2 = g_{ij}(I|\theta)dI^i dI^j, \quad (3)$$

which characterizes the statistical distance between two close sets of continuum stochastic variables  $I$  and  $I + dI$  for fixed values of control parameters  $\theta$ . This geometrical approach, hereafter referred to as *fluctuation geometry*, is inspired on the existence of a notable analogy between the general theorems of inference theory and some general fluctuation theorems recently proposed in the framework of equilibrium classical fluctuation theory [15, 16, 17, 18]. The main consequence derived from this analysis is the possibility to rephrase the original parametric family of distribution functions (2) in terms of purely geometry notions. This connection enables the direct application of powerful tools of differential geometry for the analysis of their *absolute statistical properties*§.

The paper is organized as follows. For the sake of self-consistence, the next section will be devoted to discuss the main motivation of the present proposal: the analogy between inference theory and fluctuation theory. Afterwards, it will be developed an axiomatic formulation of fluctuation geometry. Firstly, the postulates of fluctuation geometry are presented in section 3. Then, their consequences are considered in section 4 to perform a geometric reinterpretation of the statistical description. Section 5 is devoted to discuss two simple application examples of fluctuation geometry. Implications of the present approach in some statistics and physics problems will be analyzed in section 6, as example, a reconsideration of information entropy for a continuous distribution and the development of a Riemannian reformulation of Einstein fluctuation theory. Final remarks and open problems will be summarized in section 7.

## 2. Motivation

Let us start from the parametric family of distribution functions (2), which describes the behavior of a set of continuous stochastic variables  $I$  driven by a set  $\theta$  of control parameters. Let us denote by  $\mathcal{M}_\theta$  the statistical manifold constituted by all admissible values of the stochastic variables  $I$  that are accessible for a given value  $\theta$  of control parameters, which is hereafter assumed as a simply connected domain. Moreover, let us denote by  $\mathcal{P}$  the statistical manifold constituted by all admissible values of control parameters  $\theta$  (each point  $\theta \in \mathcal{P}$  represents a given distribution function). The parametric family of distribution functions (2) can be analyzed from two different perspectives:

- To study the *fluctuating behavior* of stochastic variables  $I \in \mathcal{M}_\theta$ , which is the main interest of *fluctuation theory* [17];
- To study the relationship between this fluctuation behavior and the *external control* described by the parameters  $\theta \in \mathcal{P}$ , which is the interest of *inference theory* [2].

§ Tensorial formalism of Riemannian geometry allows to study the *absolute geometric properties* of a manifold  $\mathcal{M}$  using any of its coordinate representations  $\mathcal{R}_I$ . A relevant example of absolute property is the *curvature* of the manifold  $\mathcal{M}$ , which is manifested in any coordinate representation  $\mathcal{R}_I$ . In general relativity theory, the curvature of space-time  $\mathbb{M}^4$  is identified with gravitation interaction. The effects of gravitation are *absolute* or *irreducible*, while the effects associated with the *inertial forces* are *reducible*. In fact, the existence of inertial forces crucially depends on the reference frame, that is, the specific coordinate representation of the space-time  $\mathbb{M}^4$ .

### 2.1. Fluctuation theory

Let us admit that the probability density  $\rho(I|\theta)$  is everywhere finite and differentiable, and obeys the following conditions for every point  $I_b$  located on the boundary  $\partial\mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$ ,  $I_b \in \partial\mathcal{M}_\theta$ :

$$\lim_{I \rightarrow I_b} \rho(I|\theta) = \lim_{I \rightarrow I_b} \frac{\partial}{\partial I^i} \rho(I|\theta) = 0. \quad (4)$$

The probability density  $\rho(I|\theta)$  can be considered to introduce the *differential forces*  $\eta_i(I|\theta)$ :

$$\eta_i(I|\theta) = -\frac{\partial}{\partial I^i} \log \rho(I|\theta). \quad (5)$$

By definition, the differential forces  $\eta_i(I|\theta)$  vanish in those points where the probability density  $\rho(I|\theta)$  exhibits its local maxima or its local minima. The global (local) maximum of the probability density can be regarded as a *stable (metastable) equilibrium points*  $\bar{I}$ , which can be obtained from the following *stationary and stability conditions*:

$$-\frac{\partial}{\partial I^i} \log \rho(\bar{I}|\theta) = 0, \quad -\frac{\partial^2}{\partial I^i \partial I^j} \log \rho(\bar{I}|\theta) \succ 0, \quad (6)$$

where  $A_{ij} \succ 0$  denotes that the matrix  $A_{ij}$  is positive definite. In general, the differential forces  $\eta_i(I|\theta)$  characterize the deviation of a given point  $I \in \mathcal{M}_\theta$  from these local equilibrium points. Analogously, it is convenient to introduce the *response matrix*  $\chi_{ij}(I|\theta)$ :

$$\chi_{ij}(I|\theta) = \partial_j \eta_i(I|\theta), \quad (7)$$

where  $\partial_i A = \partial A / \partial I^i$ , which describes the response of differential forces  $\eta_i(I|\theta)$  under an infinitesimal change of the variable  $I^j$ .

As stochastic variables, the expectation values of the differential forces  $\eta_i = \eta_i(I|\theta)$  identically vanish:

$$\langle \eta_i \rangle = 0, \quad (8)$$

and these quantities also obey the *fundamental and the associated fluctuation theorems* [17]:

$$\langle \eta_i \delta I^j \rangle = \delta_i^j, \quad (9)$$

$$\langle \chi_{ij} \rangle = \langle \eta_i \eta_j \rangle, \quad (10)$$

where  $\delta_i^j$  is the Kronecker delta. The previous theorems are derived from the following identity:

$$\langle \partial_i A(I|\theta) \rangle = \langle \eta_i(I|\theta) A(I|\theta) \rangle \quad (11)$$

substituting the cases  $A(I|\theta) = 1$ ,  $I^i$  and  $\eta_i$ , respectively. Here,  $A(I)$  is a differentiable function defined on the continuous variables  $I$  with definite expectation values  $\langle \partial A(I|\theta) / \partial I^i \rangle$  that obeys the following the boundary condition:

$$\lim_{I \rightarrow I_b} A(I) \rho(I|\theta) = 0. \quad (12)$$

Moreover, equation (11) follows from the integral expression:

$$\int_{\mathcal{M}_\theta} \frac{\partial v^j(I|\theta)}{\partial I^j} \rho(I|\theta) dI = \oint_{\partial\mathcal{M}_\theta} \rho(I|\theta) v^j(I|\theta) \cdot d\Sigma_j - \int_{\mathcal{M}_\theta} v^j(I|\theta) \frac{\partial \rho(I|\theta)}{\partial I^j} dI \quad (13)$$

derived from the *intrinsic exterior calculus* of the statistical manifold  $\mathcal{M}_\theta$  and the imposition of the constraint  $v^j(I|\theta) \equiv \delta_i^j A(I|\theta)$ . It is easy to realize that the identity (8) and the associated fluctuation theorem (10) are just the *stationary and stability equilibrium conditions* (6) written in

term of statistical expectation values, respectively. In particular, the positive definite character of the self-correlation matrix  $M_{ij}(\theta) = \langle \eta_i(I|\theta) \eta_j(I|\theta) \rangle$  implies the positive definition of the matrix:

$$\langle \partial_i \eta_j(I|\theta) \rangle = \left\langle -\frac{\partial^2 \log \rho(I|\theta)}{\partial I^i \partial I^j} \right\rangle \succ 0, \quad (14)$$

Remarkably, the fundamental fluctuation theorem (9) suggests the *statistical complementarity* of the variable  $I^i$  and its conjugated differential force  $\eta_i = \eta_i(I|\theta)$ . Using the Schwartz inequality  $\langle \delta A \delta B \rangle^2 \leq \langle \delta A^2 \rangle \langle \delta B^2 \rangle$ , one obtains the following inequality:

$$\Delta I^i \Delta \eta_i \geq 1, \quad (15)$$

where  $\Delta x = \sqrt{\langle \delta x^2 \rangle}$  is the *statistical uncertainty* of the quantity  $x$ . Clearly, this last inequality exhibits the same mathematical appearance of *Heisenberg's uncertainty relation*  $\Delta q \Delta p \geq \hbar$ . Recently, this result was employed to show the existence of uncertainty relations involving conjugated thermodynamic quantities [15, 16]. Equation (15) can be generalized considering the inverse  $M^{ij}(\theta)$  of the self-correlation matrix of the differential forces  $M_{ij}(\theta) = \langle \eta_i(I|\theta) \eta_j(I|\theta) \rangle$ . Denoting by  $C^{ij}(\theta) = \langle \delta I^i \delta I^j \rangle$  the self-correlation matrix of the stochastic variables  $I$ , it is possible to obtain the following matrical inequalities:

$$C^{ij}(\theta) - M^{ij}(\theta) \succeq 0. \quad (16)$$

This last inequality is directly obtained from the positive definition of the self-correlation matrix  $K^{ij}(\theta) = \langle J^i(I|\theta) J^j(I|\theta) \rangle$  of the auxiliary quantities  $J^i(I|\theta) = \delta I^i + M^{ij}(\theta) \eta_j(I|\theta)$ . Accordingly, the self-correlation matrix  $C^{ij}(\theta)$  of stochastic variables  $I$  is inferior bound by the inverse  $M^{ij}(\theta)$  of the self-correlation matrix of the differential forces  $\eta_i$ .

## 2.2. Inference theory

Inference theory can be described as the problem of deciding how well a set of outcomes  $\mathcal{I} = \{I^{(1)}, I^{(2)}, \dots, I^{(m)}\}$  obtained from independent measurements fits a proposed distribution function  $dp(I|\theta)$  [2]. This question is fully equivalent to infer the values of control parameters  $\theta$  from this last experimental information. To make inferences about control parameters, one employs *estimators*  $\hat{\theta}^\alpha = \theta^\alpha(\mathcal{I})$ , that is, functions on the outcomes  $\mathcal{I} \in \mathcal{M}_\theta^m$ , where  $\mathcal{M}_\theta^m = \mathcal{M}_\theta \otimes \mathcal{M}_\theta \dots \otimes \mathcal{M}_\theta$  (m-times the external product of the statistical manifold  $\mathcal{M}_\theta$ ). The values of these functions pretend to be the best guess for  $\theta^\alpha$ .

Let us admit that the probability density  $\rho(I|\theta)$  is everywhere differentiable and finite on the statistical manifold  $\mathcal{P}$  of control parameters  $\theta$ . Let us start introducing the statistical expectation values  $\langle A(\mathcal{I}|\theta) \rangle$  as follows:

$$\langle A(\mathcal{I}|\theta) \rangle = \int_{\mathcal{M}_\theta^m} A(\mathcal{I}|\theta) \varrho(\mathcal{I}|\theta) d\mathcal{I}, \quad (17)$$

where  $d\mathcal{I} = dI^{(1)} dI^{(2)} \dots dI^{(m)}$  and  $\varrho(\mathcal{I}|\theta)$  is the so-called *likelihood function*:

$$\varrho(\mathcal{I}|\theta) = \prod_{i=1}^m \rho(I^{(i)}|\theta). \quad (18)$$

Taking the partial derivative  $\partial_\alpha = \partial/\partial\theta^\alpha$  of Eq.(17), one obtains the following mathematical identity:

$$\langle \partial_\alpha A(\mathcal{I}|\theta) \rangle - \partial_\alpha \langle A(\mathcal{I}|\theta) \rangle = \langle A(\mathcal{I}|\theta) v_\alpha(\mathcal{I}|\theta) \rangle, \quad (19)$$

where:

$$\hat{v}_\alpha = v_\alpha(\mathcal{I}|\theta) = -\frac{\partial}{\partial\theta^\alpha} \log \rho(\mathcal{I}|\theta) \quad (20)$$

are the components of the *score vector*  $\hat{v} = \{\hat{v}_\alpha\}$ . Substituting  $A(\mathcal{I}|\theta) = 1$  into Eq.(19), one arrives at the vanishing of the expectation values of the score vector components:

$$\langle v_\alpha(\mathcal{I}|\theta) \rangle = 0. \quad (21)$$

Let us consider now any *unbiased estimator*  $A(\mathcal{I}|\theta) = \theta^\alpha(\mathcal{I})$  of the parameter  $\theta^\alpha$ ,  $\langle \theta^\alpha(\mathcal{I}) \rangle = \theta^\alpha$ , as well as the score vector component  $A(\mathcal{I}|\theta) = v_\beta(\mathcal{I}|\theta)$ . Substituting these quantities into identity (19), it is possible to obtain the following results:

$$\langle \delta\theta^\alpha(\mathcal{I}) v_\beta(\mathcal{I}|\theta) \rangle = -\delta_\beta^\alpha, \quad (22)$$

$$\langle \partial_\beta v_\alpha(\mathcal{I}|\theta) \rangle = \langle v_\alpha(\mathcal{I}|\theta) v_\beta(\mathcal{I}|\theta) \rangle. \quad (23)$$

It is easy to realize that the identities (21) and (23) can be regarded as the stationary and stability conditions of the known method of *maximum likelihood estimators* [2]. According to this method, the best values of the parameters  $\theta$  should maximize the logarithm of the *likelihood function*  $\varrho(\mathcal{I}|\theta)$  for a given set of outcomes  $\mathcal{I}$ . Such an exigence leads to the following *stationary and stability conditions*:

$$-\frac{\partial}{\partial\theta^\alpha} \log \varrho(\mathcal{I}|\bar{\theta}) = 0, -\frac{\partial^2}{\partial\theta^\alpha \partial\theta^\beta} \log \varrho(\mathcal{I}|\bar{\theta}) \succ 0, \quad (24)$$

which should be solved to obtain the *maximum likelihood estimators*  $\hat{\theta}_{mle}^\alpha = \bar{\theta}^\alpha(\mathcal{I})$ . On the other hand, the identity (22) also suggests the statistical complementarity between the estimator  $\hat{\theta}^\alpha = \theta^\alpha(\mathcal{I})$  and its conjugated score vector component  $v_\alpha$ . Using the Schwartz inequality, one obtains the following uncertainty-like inequality:

$$\Delta\hat{\theta}^\alpha \Delta v_\alpha \geq 1. \quad (25)$$

This result can be easily improved introducing the inverse matrix  $g^{\alpha\beta}(\theta)$  of the self-correlation matrix  $g_{\alpha\beta}(\theta) = \langle v_\alpha(\mathcal{I}|\theta) v_\beta(\mathcal{I}|\theta) \rangle$  and the auxiliary quantity  $X^\alpha = \delta\hat{\theta}^\alpha - g^{\alpha\beta} v_\beta$ . Thus, one can compose the positive definite form:

$$\langle (\lambda_\alpha X^\alpha)^2 \rangle = \langle X^\alpha X^\beta \rangle \lambda_\alpha \lambda_\beta \geq 0, \quad (26)$$

which leads to the positive definition of the matrix:

$$\langle \delta\hat{\theta}^\alpha \delta\hat{\theta}^\beta \rangle - g^{\alpha\beta}(\theta) \succeq 0. \quad (27)$$

This last inequality is the famous *Cramer-Rao theorem* of inference theory [2, 3] that imposes an inferior bound to the efficiency of unbiased estimators  $\hat{\theta}^\alpha$ , where the self-correlation matrix  $g_{\alpha\beta}(\theta)$ :

$$g_{\alpha\beta}(\theta) = \langle v_\alpha(\mathcal{I}|\theta) v_\beta(\mathcal{I}|\theta) \rangle \quad (28)$$

is the Fisher's information matrix referred to in the introductory section.

Inference theory	Fluctuation theory
$v(\mathcal{I} \theta) = -\nabla_\theta \log \rho(\mathcal{I} \theta)$	$\eta(I \theta) = -\nabla_I \log \rho(I \theta)$
$\langle v(\mathcal{I} \theta) \rangle = 0$	$\langle \eta(I \theta) \rangle = 0$
$\langle v(\mathcal{I} \theta) \cdot \delta \hat{\theta} \rangle = -\mathbf{1}_\theta$	$\langle \eta(I \theta) \cdot \delta \mathbf{I} \rangle = \mathbf{1}_I$
$\langle \nabla_\theta \cdot v(\mathcal{I} \theta) \rangle = \langle v(\mathcal{I} \theta) \cdot v(\mathcal{I} \theta) \rangle$	$\langle \nabla_I \cdot \eta(I \theta) \rangle = \langle \eta(I \theta) \cdot \eta(I \theta) \rangle$

**Table 1.** Analogy between inference theory and fluctuation theory.

### 2.3. Analogy between inference theory and fluctuation theory

Fluctuation theory and inference theory provide two different but *complementary* characterizations for a given parametric family of distribution functions  $dp(I|\theta)$ . Formally, these two statistical frameworks can be regarded as *dual counterpart approaches* because of the great analogy among their main definitions and theorems, as clearly evidenced in table 1. To simplify the notation, it was introduced here the gradient operators  $\partial_i \rightarrow \nabla_I$  and  $\partial_\alpha \rightarrow \nabla_\theta$ , the diadic products  $\mathbf{A} \cdot \mathbf{B} = A_i B_j \mathbf{e}^i \cdot \mathbf{e}^j$  and  $\xi \cdot \psi = \xi_\alpha \psi_\beta \epsilon^\alpha \cdot \epsilon^\beta$  and the Kronecker delta  $\delta_j^i \rightarrow \mathbf{1}_I$  and  $\delta_\beta^\alpha \rightarrow \mathbf{1}_\theta$ .

Remarkably, the analogy between fluctuation theory and inference theory is incomplete in regard to their respective *geometric features*. The parametric family  $dp(I|\theta)$  is expressed in the representations  $\mathcal{R}_I$  and  $\mathcal{R}_\theta$  of the statistical manifolds  $\mathcal{M}_\theta$  and  $\mathcal{P}$ , respectively. Equivalently, the same parametric family can be also rewritten using the representations  $\mathcal{R}_\Theta$  and  $\mathcal{R}_\nu$  of the manifolds  $\mathcal{M}_\theta$  and  $\mathcal{P}$ , which implies the consideration of the coordinate changes  $\Theta(I) : \mathcal{R}_I \rightarrow \mathcal{R}_\Theta$  and  $\nu(\theta) : \mathcal{R}_\theta \rightarrow \mathcal{R}_\nu$ . Under these parametric changes, the Fisher's inference matrix (28) behaves as the components of a second rank covariant tensor:

$$g_{\gamma\delta}(\nu) = \frac{\partial \theta^\alpha}{\partial \nu^\gamma} \frac{\partial \theta^\beta}{\partial \nu^\delta} g_{\alpha\beta}(\theta). \quad (29)$$

The existence of these last transformation rules guarantees the invariance of the inference distance notion (1). Thus, the statistical manifold  $\mathcal{P}$  of control parameters  $\theta$  can be endowed of a *Riemannian structure*. The relevance of the distance notion (1) can be understood considering the asymptotic expression of the distribution function of the *efficient unbiased estimators*  $\hat{\theta}_{eff}(\mathcal{I})$ :

$$dQ^m(\vartheta|\theta) = \int_{\mathcal{M}_\theta^m} \delta \left[ \vartheta - \hat{\theta}_{eff}(\mathcal{I}) \right] \varrho(\mathcal{I}|\theta) d\mathcal{I} \quad (30)$$

when the number of outcomes  $m$  is sufficiently large. Since  $g_{\alpha\beta}(\theta) \propto m$ , one can obtain the following approximation formula:

$$dQ^m(\vartheta|\theta) \simeq \exp \left[ -\frac{1}{2} g_{\alpha\beta}(\theta) \Delta \vartheta^\alpha \Delta \vartheta^\beta \right] \sqrt{\left| \frac{g_{\alpha\beta}(\theta)}{2\pi} \right|} d\vartheta, \quad (31)$$

where  $\Delta \vartheta^\alpha = \vartheta^\alpha - \theta^\alpha$ . Accordingly, the distance notion (1) provides the *distinguishing probability* between two close distribution functions of the parametric family  $dp(I|\theta)$  through the inferential procedure.

The analogy between fluctuation theory and inference theory strongly suggests the existence of a *counterpart approach of inference geometry in the framework of fluctuation theory*, that is, the existence of a Riemannian distance notion (3) to characterize the statistical separation between close points  $I$  and  $I + dI \in \mathcal{M}_\theta$ . Unfortunately, the underlying analogy is insufficient to introduce the particular expression of the metric tensor  $g_{ij}(I|\theta)$ . For example, it is easy to check that the

fluctuation theorems (8)-(10) can be also expressed in the new coordinate representation  $\mathcal{R}_\Theta$  [18], as example, the associated fluctuation theorem:

$$\langle \partial_i \eta_j(\Theta|\theta) \rangle = \langle \eta_i(\Theta|\theta) \eta_j(\Theta|\theta) \rangle, \quad (32)$$

where  $\partial_i = \partial/\partial\Theta^i$  and  $\eta_i(\Theta|\theta)$ :

$$\eta_i(\Theta|\theta) = -\frac{\partial}{\partial\Theta^i} \log \rho(\Theta|\theta). \quad (33)$$

However, the self-correlation matrix  $\tilde{M}_{ij}(\theta) = \langle \eta_i(\Theta|\theta) \eta_j(\Theta|\theta) \rangle$  associated with the new coordinate representation  $\mathcal{R}_\Theta$  is not related by *local transformation rules* to its counterpart expression  $M_{ij}(\theta) = \langle \eta_i(I|\theta) \eta_j(I|\theta) \rangle$  in the old coordinate representation  $\mathcal{R}_I$ . The self-correlation matrix of the differential forces  $M_{ij}(\theta) = \langle \eta_i(I|\theta) \eta_j(I|\theta) \rangle$  is a matrix function defined on the statistical manifold  $\mathcal{P}$  of control parameters  $\theta$ , while the metric tensor  $g_{ij}(I|\theta)$  is a *tensorial entity* defined on the statistical manifolds  $\mathcal{M}_\theta$  and  $\mathcal{P}$ . As expected, the definition of the metric tensor  $g_{ij}(I|\theta)$  cannot involve integral expressions over the manifold  $\mathcal{M}_\theta$  as the case of inference metric tensor (28).

### 3. Fluctuation geometry

Fluctuation geometry can be formulated starting from a set of axioms that combine the statistical nature of the manifold  $\mathcal{M}_\theta$  and the notions of differential geometry. These axioms specify the way to introduce the metric tensor  $g_{ij}(I|\theta)$  associated with the parametric family (2). This section is devoted to discuss these axioms and their most direct consequences.

#### 3.1. Postulates

**Axiom 1** *The manifold of the stochastic variables  $\mathcal{M}_\theta$  possesses a **Riemannian structure**, that is, it is provided of a **metric tensor**  $g_{ij}(I|\theta)$  and a **torsionless covariant differentiation**  $D_i$  that obeys the following constraints:*

$$D_k g_{ij}(I|\theta) = 0. \quad (34)$$

**Definition 1** *The Riemannian structure on the statistical manifold  $\mathcal{M}_\theta$  allows to introduce the **invariant volume element** as follows:*

$$d\mu(I|\theta) = \sqrt{\left| \frac{g_{ij}(I|\theta)}{2\pi} \right|} dI, \quad (35)$$

where  $|g_{ij}(I|\theta)|$  denotes the absolute value of the metric tensor determinant.

**Axiom 2** *There exist a differentiable scalar function  $\mathcal{S}(I|\theta)$  defined on the statistical manifold  $\mathcal{M}_\theta$ , hereafter referred to as the **information potential**, whose knowledge determines the distribution function  $dp(I|\theta)$  of the stochastic variables  $I \in \mathcal{M}_\theta$  as follows:*

$$dp(I|\theta) = \exp[\mathcal{S}(I|\theta)] d\mu(I|\theta). \quad (36)$$

**Definition 2** *Let us consider an arbitrary curve given in parametric form  $I(t) \in \mathcal{M}_\theta$  with fixed extreme points  $I(t_1) = P$  and  $I(t_2) = Q$ . Adopting the following notation:*

$$\dot{I}^i = \frac{dI^i(t)}{dt}, \quad (37)$$

the **length**  $\Delta s$  of this curve can be expressed as:

$$\Delta s = \int_{t_1}^{t_2} \sqrt{g_{ij}[I(t)|\theta] \dot{I}^i(t) \dot{I}^j(t)} dt. \quad (38)$$

**Definition 3** It is say that the curve  $I(t) \in \mathcal{M}_\theta$  exhibits a **unitary affine parametrization** when its parameter  $t$  satisfies the following constraint:

$$g_{ij}[I(t)|\theta] \dot{I}^i(t) \dot{I}^j(t) = 1. \quad (39)$$

**Definition 4** A **geodesic** is the curve  $I_g(t)$  with minimal length ( $4\theta$ ) between two fixed arbitrary points  $(P, Q) \in \mathcal{M}_\theta$ . Moreover, the **distance**  $\mathfrak{D}_\theta(P, Q)$  between these two points  $(P, Q)$  is given by the length of its associated geodesic  $I_g(t)$ :

$$\mathfrak{D}_\theta(P, Q) = \int_{t_1}^{t_2} \sqrt{g_{ij}[I_g(t)|\theta] \dot{I}_g^i(t) \dot{I}_g^j(t)} dt. \quad (40)$$

**Definition 5** Let us consider a differentiable curve  $I(t) \in \mathcal{M}_\theta$  with an unitary affine parametrization. The **information dissipation**  $\Phi(t)$  along the curve  $I(t)$  is defined as follows:

$$\Phi(t) = \frac{d\mathcal{S}[I(t)|\theta]}{dt}. \quad (41)$$

**Axiom 3** The length  $\Delta s$  of any interval  $(t_1, t_2)$  of an arbitrary geodesic  $I_g(t) \in \mathcal{M}_\theta$  with a unitary affine parametrization is given by the negative of the variation of its information dissipation  $\Delta\Phi(t)$ :

$$\Delta s = -\Delta\Phi(t) = \Phi(t_1) - \Phi(t_2). \quad (42)$$

**Axiom 4** If  $\mathcal{M}_\theta$  is not a closed manifold,  $\partial\mathcal{M}_\theta \neq \emptyset$ , the probability density  $\rho(I|\theta)$  associated with distribution function (36) vanishes with its first partial derivatives for any point on the boundary  $\partial\mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$ .

### 3.2. Analysis of axioms and their direct consequences

**Axiom 1** postulates the existence of the metric tensor  $g_{ij}(I|\theta)$  defined on the statistical manifold  $\mathcal{M}_\theta$ . Even, this axiom specifies the *Riemannian structure* of the manifold  $\mathcal{M}_\theta$  starting from the knowledge of the metric tensor  $g_{ij}(I|\theta)$ , e.g.: the *covariant differentiation*  $D_i$  and the *curvature tensor*  $R_{ijkl}(I|\theta)$ . Equation (34) is an strong constraint of Riemannian geometry that determines a natural *affine connections*  $\Gamma_{ij}^k$  for the covariant differentiation  $D_i$ , specifically, the so-called *Levi-Civita connection* [19]:

$$\Gamma_{ij}^k(I|\theta) = g^{km} \frac{1}{2} \left( \frac{\partial g_{im}}{\partial I^j} + \frac{\partial g_{jm}}{\partial I^i} - \frac{\partial g_{ij}}{\partial I^m} \right). \quad (43)$$

The knowledge of the affine connections  $\Gamma_{ij}^k$  allows the introduction of the *curvature tensor*  $R_{ijk}^l = R_{ijk}^l(I|\theta)$  of the manifold  $\mathcal{M}_\theta$ :

$$R_{ijk}^l = \frac{\partial}{\partial I^i} \Gamma_{jk}^l - \frac{\partial}{\partial I^j} \Gamma_{ik}^l + \Gamma_{im}^l \Gamma_{jk}^m - \Gamma_{jm}^l \Gamma_{ik}^m, \quad (44)$$

which is also derived from the knowledge of the metric tensor  $g_{ij}(I|\theta)$  and its first and second partial derivatives.

**Axiom 2** postulates the probabilistic nature of the manifold  $\mathcal{M}_\theta$ , in particular, the existence of the distribution function  $dp(I|\theta)$  and the information potential  $\mathcal{S}(I|\theta)$ . Equivalently, this axiom provides a formal definition for the information potential  $\mathcal{S}(I|\theta)$  when one starts from the knowledge of the parametric family (2). The probability density  $\rho(I|\theta)$  of the parametric family (2) obeys the transformation rule of a tensorial density:

$$\rho(\Theta|\theta) = \rho(I|\theta) \left| \frac{\partial \Theta}{\partial I} \right|^{-1} \quad (45)$$



under coordinate change  $\Theta(I) : \mathcal{R}_I \rightarrow \mathcal{R}_\Theta$  of the statistical manifold  $\mathcal{M}_\theta$ . The covariance of the metric tensor  $g_{ij}(I|\theta)$ :

$$g_{ij}(\Theta|\theta) = \frac{\partial I^m}{\partial \Theta^i} \frac{\partial I^n}{\partial \Theta^j} g_{mn}(I|\theta) \quad (46)$$

implies that the pre-factor of the invariant volume element (35) also behaves as a tensorial density:

$$\sqrt{\left| \frac{g_{ij}(\Theta|\theta)}{2\pi} \right|} = \sqrt{\left| \frac{g_{ij}(I|\theta)}{2\pi} \right|} \left| \frac{\partial \Theta}{\partial I} \right|^{-1}. \quad (47)$$

Admitting that the metric tensor determinant  $|g_{ij}(I|\theta)|$  is non-vanishing everywhere, it is possible to introduce *probability weight*  $\omega(I|\theta)$ :

$$\omega(I|\theta) = \rho(I|\theta) \sqrt{|2\pi g^{ij}(I|\theta)|}, \quad (48)$$

which represents a *scalar function* defined on the manifold  $\mathcal{M}_\theta$ . Since the statistical manifold  $\mathcal{M}_\theta$  possesses a Riemannian structure, the integration over the usual volume element  $dI$  can be replaced by the invariant volume element  $d\mu(I|\theta)$ . This consideration allows to rephrase the parametric family (2) in the following equivalent representation:

$$dp(I|\theta) = \omega(I|\theta) d\mu(I|\theta), \quad (49)$$

which explicitly exhibits the covariance of the distribution function under the coordinate reparametrizations of the manifold  $\mathcal{M}_\theta$ . The information potential  $\mathcal{S}(I|\theta)$  is defined by the logarithm of the probability weight  $\omega(I|\theta)$ :

$$\mathcal{S}(I|\theta) = \log \omega(I|\theta), \quad (50)$$

which also represents a scalar function defined on the statistical manifold  $\mathcal{M}_\theta$ . As discussed in section 6, the negative of the information potential  $\mathcal{S}(I|\theta)$  can be regarded as a *local invariant measure of the information content* in the framework of information theory. Additionally,  $\mathcal{S}(I|\theta)$  can be identified with the *scalar entropy of a closed system* in the framework of classical fluctuation theory [18]. Given the probability density  $\rho(I|\theta)$ , the information potential  $\mathcal{S}(I|\theta)$  depends on the metric tensor  $g_{ij}(I|\theta)$  of the statistical manifold  $\mathcal{M}_\theta$ . **Axiom 1** postulates the existence of this tensor, but its specific definition is still arbitrary. **Axiom 3** eliminates such an ambiguity. In fact, it establishes a direct connection between the distance notion (3) and the information dissipation (41), or equivalently, between the metric tensor  $g_{ij}(I|\theta)$  and the information potential  $\mathcal{S}(I|\theta)$ .

**Theorem 1** *The metric tensor  $g_{ij}(I|\theta)$  can be identified with the negative of the **covariant Hessian**  $\mathcal{H}_{ij}(I|\theta)$  of the information potential  $\mathcal{S}(I|\theta)$ :*

$$g_{ij}(I|\theta) = -\mathcal{H}_{ij}(I|\theta) = -D_i D_j \mathcal{S}(I|\theta). \quad (51)$$

**Corollary 1** *The information potential  $\mathcal{S}(I|\theta)$  is locally concave everywhere and the metric tensor  $g_{ij}(I|\theta)$  is positive definite on the statistical manifold  $\mathcal{M}_\theta$ .*

**Proof.** The searching of the curve with minimal length (40) between two arbitrary points  $(P, Q)$  is a variational problem that leads to the so-called *geodesic differential equations* [19]:

$$\dot{I}_g^k(t) D_k \dot{I}_g^i(t) = \ddot{I}_g^i(t) + \Gamma_{mn}^i [I_g(t)|\theta] \dot{I}_g^m(t) \dot{I}_g^n(t) = 0, \quad (52)$$

which describes the geodesics  $I_g(t)$  with a unitary affine parametrization. Equations (41) and (42) can be rephrased as follows:

$$\Delta s = -\Delta \Phi(s) \rightarrow \frac{d\Phi(s)}{ds} = \frac{d^2 \mathcal{S}}{ds^2} = -1. \quad (53)$$

Considering the geodesic differential equations (52), the constraint (53) is rewritten as:

$$\frac{d^2 \mathcal{S}}{ds^2} = \ddot{I}_g^k \frac{\partial \mathcal{S}}{\partial I^k} + \dot{I}_g^i \dot{I}_g^j \frac{\partial^2 \mathcal{S}}{\partial I^i \partial I^j} = \dot{I}_g^i \dot{I}_g^j \left\{ \frac{\partial^2 \mathcal{S}}{\partial I^i \partial I^j} - \Gamma_{ij}^k \frac{\partial \mathcal{S}}{\partial I^k} \right\}, \quad (54)$$

which involves the *covariant Hessian*  $\mathcal{H}_{ij}$ :

$$\mathcal{H}_{ij} = D_i D_j \mathcal{S} = \frac{\partial^2 \mathcal{S}}{\partial I^i \partial I^j} - \Gamma_{ij}^k \frac{\partial \mathcal{S}}{\partial I^k}. \quad (55)$$

Combining equations (54)-(55) and the constraint (39), one obtains the following expression:

$$(g_{ij} + \mathcal{H}_{ij}) \dot{I}_g^i \dot{I}_g^j = 0. \quad (56)$$

Its covariant character leads to Eq.(51). **Corollary 1**, that is, the concave character of the information potential  $\mathcal{S}(I|\theta)$  and the positive definition of the metric tensor  $g_{ij}(I|\theta)$  are direct consequences of equation (53). ■

**Corollary 2** *The metric tensor  $g_{ij} = g_{ij}(I|\theta)$  can be obtained from the probability density  $\rho = \rho(I|\theta)$  through the following set of covariant second-order partial differential equations:*

$$g_{ij} = -\frac{\partial^2 \log \rho}{\partial I^i \partial I^j} + \Gamma_{ij}^k \frac{\partial \log \rho}{\partial I^k} + \frac{\partial}{\partial I^i} \Gamma_{jk}^k - \Gamma_{ij}^k \Gamma_{kl}^l. \quad (57)$$

*The admissible solutions for the metric tensor  $g_{ij}$  should be finite and differentiable everywhere, including also, the boundary  $\partial \mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$ .*

**Proof.** Expression (57) is derived from equation (51) rewriting the information potential as  $\mathcal{S}(I|\theta) \equiv \log \rho(I|\theta) - \log \sqrt{|g_{ij}(I|\theta)|/2\pi}$  and considering the following identity:

$$\Gamma_{ij}^j(I|\theta) \equiv \frac{\partial \log \sqrt{|g_{ij}(I|\theta)|/2\pi}}{\partial I^i}, \quad (58)$$

which is a known property of the Levi-Civita connection (43). ■

**Axiom 4** talks about the asymptotic behavior of the distribution function (36) for any point  $I_b$  on the boundary  $\partial \mathcal{M}_\theta$ :

$$\lim_{I \rightarrow I_b} \rho(I|\theta) = \lim_{I \rightarrow I_b} \frac{\partial}{\partial I^i} \rho(I|\theta) = 0. \quad (59)$$

These last conditions are necessary to obtain the general fluctuation theorems (8)-(10) reviewed in the previous section. Moreover, this axiom will be employed to analyze the character of stationary points (maxima and minima) of the information potential  $\mathcal{S}(I|\theta)$ .

**Remark 1** *The boundary conditions (59) are independent from the admissible coordinate representation  $\mathcal{R}_I$  of the statistical manifold  $\mathcal{M}_\theta$ . Moreover, the probability weight  $\omega(I|\theta)$  vanishes on the boundary  $\partial \mathcal{M}_\theta$  of the manifold  $\mathcal{M}_\theta$ .*

**Proof.** This remark is a direct consequence of the transformation rule of the probability density (45) as well as the ones associated with its partial derivatives:

$$\frac{\partial \rho(\Theta|\theta)}{\partial \Theta^i} = \frac{\partial I^j}{\partial \Theta^i} \left\{ \frac{\partial \rho(I|\theta)}{\partial I^j} - \rho(I|\theta) \frac{\partial}{\partial I^j} \log \left| \frac{\partial \Theta}{\partial I} \right| \right\} \left| \frac{\partial \Theta}{\partial I} \right|^{-1} \quad (60)$$

under a coordinate change  $\Theta(I) : \mathcal{R}_I \rightarrow \mathcal{R}_\Theta$  with Jacobian  $|\partial \Theta / \partial I|$  finite and differentiable everywhere. Since the metric tensor determinant  $|g_{ij}(I|\theta)|$  is non-vanishing everywhere, **Axiom 4** directly implies the vanishing of the probability weight  $\omega(I|\theta)$  on the boundary  $\partial \mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$ . ■

#### 4. Geometric reinterpretation of the statistical description

The question about the *existence and uniqueness* of solutions obtained from the problem (57) cannot be fully analyzed in this work because of its complexity. This section is devoted to discuss some consequences derived from the existence of a given particular solution  $g_{ij}(I|\theta)$ .

##### 4.1. Gaussian representation

**Definition 6** The covariant components of the **gradiental vector**  $\psi_i(I|\theta)$  are defined from the information potential  $\mathcal{S}(I|\theta)$  as follows:

$$\psi_i(I|\theta) = -D_i \mathcal{S}(I|\theta) \equiv -\partial \mathcal{S}(I|\theta) / \partial I^i. \quad (61)$$

Using the metric tensor  $g^{ij}(I|\theta)$ , it is possible to obtain its contravariant counterpart  $\psi^i(I|\theta)$ :

$$\psi^i(I|\theta) = g^{ij}(I|\theta) \psi_j(I|\theta), \quad (62)$$

as well as its the square norm  $\psi^2 = \psi^2(I|\theta)$ :

$$\psi^2(I|\theta) = \psi^i(I|\theta) \psi_i(I|\theta). \quad (63)$$

**Theorem 2** The information potential  $\mathcal{S}(I|\theta)$  can be expressed in terms of the square norm of the gradiental vector as follows:

$$\mathcal{S}(I|\theta) = \mathcal{P}(\theta) - \frac{1}{2} \psi^2(I|\theta), \quad (64)$$

where  $\mathcal{P}(\theta)$  is a certain function on control parameters  $\theta$ , which is hereafter referred to as the **gaussian potential**.

**Proof.** Let us introduce the scalar function  $\mathcal{P}(I|\theta)$ :

$$\mathcal{P}(I|\theta) = \mathcal{S}(I|\theta) + \frac{1}{2} g^{ij}(I|\theta) \psi_i(I|\theta) \psi_j(I|\theta). \quad (65)$$

It is easy to verify that its covariant derivatives:

$$\begin{aligned} D_k \mathcal{P}(I|\theta) &= D_k \mathcal{S}(I|\theta) + \frac{1}{2} \{ \psi_i(I|\theta) \psi_j(I|\theta) D_k g^{ij}(I|\theta) + \\ &+ g^{ij}(I|\theta) [\psi_i(I|\theta) D_k \psi_j(I|\theta) + \psi_j(I|\theta) D_k \psi_i(I|\theta)] \} \end{aligned} \quad (66)$$

vanish as direct consequences of the metric tensor properties (34) and (51), as well as definition (61) of the gradiental vector  $\psi_i(I|\theta)$ . Since the covariant derivatives of any scalar function are given by the usual partial derivatives:

$$D_k \mathcal{P}(I|\theta) = \frac{\partial}{\partial I^k} \mathcal{P}(I|\theta) = 0, \quad (67)$$

the scalar function  $\mathcal{P}(I|\theta)$  can only depend on the control parameters  $\theta$ :

$$\mathcal{P}(I|\theta) \equiv \mathcal{P}(\theta). \quad (68)$$

Mathematically speaking, the scalar function (65) can be regarded as a first integral of the set of covariant differential equations (57). ■

**Corollary 3** The value of information potential  $\mathcal{S}(I|\theta)$  at all its extreme points derived from the stationary condition:

$$\psi^2(\bar{I}|\theta) = 0 \quad (69)$$

is exactly given by the gaussian potential  $\mathcal{P}(\theta)$ .

**Corollary 4** *The distribution function (36) admits the following **gaussian representation**:*

$$dp(I|\theta) = \frac{1}{\mathcal{Z}(\theta)} \exp \left[ -\frac{1}{2} \psi^2(I|\theta) \right] d\mu(I|\theta). \quad (70)$$

Here, the factor  $\mathcal{Z}(\theta)$  is related to the gaussian potential  $\mathcal{P}(\theta)$  as follows:

$$\mathcal{P}(\theta) = -\log \mathcal{Z}(\theta), \quad (71)$$

which is hereafter referred to as the **gaussian partition function**.

#### 4.2. Maximum and completeness theorems

**Theorem 3** *The information potential  $\mathcal{S}(I|\theta)$  exhibits a unique stationary point  $\bar{I}$  in the statistical manifold  $\mathcal{M}_\theta$ , which corresponds to its global maximum.*

**Proof.** The information potential  $\mathcal{S}(I|\theta)$  should exhibit at least a stationary point  $\bar{I}$  where takes place the stationary condition (69). This conclusion follows from the vanishing of the scalar weight of distribution function:

$$\omega(I|\theta) = \exp [\mathcal{S}(I|\theta)] \quad (72)$$

on the boundary  $\partial\mathcal{M}_\theta$ , as well as its character nonnegative, finite and differentiable on the simply connected manifold  $\mathcal{M}_\theta$ . Since the information potential  $\mathcal{S}(I|\theta)$  is a concave function, its stationary points can only correspond to local maxima. Let us suppose the existence of at least two stationary points  $\bar{I}_1$  and  $\bar{I}_2$  as well as the geodesic  $I_g(t)$  that connects these points. According to constraint (53), the information dissipation  $\Phi(t)$  is a monotonous function along the curve  $I_g(t)$ . Therefore,  $\Phi(t)$  should exhibit different values at the stationary points  $\bar{I}_1$  and  $\bar{I}_2$ , which is absurdum since the information dissipation  $\Phi(t)$  identically vanishes for any stationary point of the information potential  $\mathcal{S}(I|\theta)$ :

$$\Phi(t) = -\dot{I}^i(t) \psi_i[I(t)|\theta]. \quad (73)$$

Consequently, there exist only one stationary point that corresponds with the global maximum of the information potential  $\mathcal{S}(I|\theta)$ . ■

**Theorem 4** *Any hyper-surface of constant information potential  $\mathcal{S}(I|\theta)$  is just the boundary of a  $n$ -dimensional sphere  $S^n(\bar{I}, \ell) \subset \mathcal{M}_\theta$  centered at the point  $\bar{I}$  with global maximum information potential, where  $n$  is the dimension of the manifold  $\mathcal{M}_\theta$ . Moreover, the information potential  $\mathcal{S}$  depends on the radius  $\ell$  of this  $n$ -dimensional sphere as follows:*

$$\mathcal{S} = \mathcal{P}(\theta) - \frac{1}{2} \ell^2. \quad (74)$$

**Proof.** By definition, the vector field  $v^i(I|\theta)$ :

$$v^i(I|\theta) = \frac{\psi^i(I|\theta)}{\psi(I|\theta)} \quad (75)$$

is the unitary normal vector of the hyper-surface with constant information potential  $\mathcal{S}(I|\theta)$ . It is easy to verify that the vector field  $v^i(I|\theta)$  obeys the geodesic equations (52):

$$v^k(I|\theta) D_k v^i(I|\theta) = \frac{v^k(I|\theta)}{\psi(I|\theta)} [\delta_k^i - v^i(I|\theta) v_k(I|\theta)] = 0. \quad (76)$$

Hence,  $v^i(I|\theta)$  can be regarded as the tangent vector:

$$\frac{dI_g^i(s|\mathbf{e})}{ds} = v^i[I_g(s|\mathbf{e})|\theta] \quad (77)$$

of geodesic family  $I_g(s|\mathbf{e})$  with unitary affine parametrization centered at the point  $\bar{I}$  with maximum information potential  $\mathcal{S}(I|\theta)$ ,  $I_g(s=0|\mathbf{e}) = \bar{I}$ . Moreover, the constant unitary vector  $\mathbf{e}$  parameterizes geodesics with different directions at the origin,  $\dot{I}_g(s=0|\mathbf{e}) = \mathbf{e}$ . The information dissipation  $\Phi(s|\mathbf{e})$  along any of these geodesics is given by the negative of the norm of the gradiental vector:

$$\Phi(s|\mathbf{e}) = -\frac{dI_g^i(s|\mathbf{e})}{ds} \psi_i[I_g(s|\mathbf{e})|\theta] = -\psi[I_g(s|\mathbf{e})|\theta]. \quad (78)$$

Considering equation (42), the norm  $\psi(I|\theta)$  can be related to the length  $\Delta s$  of the geodesic that connects an arbitrary point  $I$  with the point  $\bar{I}$  with maximum information potential, that is, the distance  $\mathfrak{D}_\theta(I, \bar{I})$  between the points  $I$  and  $\bar{I}$ :

$$\psi(I|\theta) = \mathfrak{D}_\theta(I, \bar{I}). \quad (79)$$

According to the gaussian decomposition (64), the hyper-surface with constant information potential  $\mathcal{S}(I|\theta)$  is also the hyper-surface where the norm of gradiental generalized forces  $\psi(I|\theta)$  is kept constant, that is, the boundary of a  $n$ -dimensional sphere  $S^n(\bar{I}, \ell)$  centered at the point  $\bar{I}$  with maximum information potential. ■

**Corollary 5** *The distribution function (36) can be expressed in the following **Riemannian gaussian representation**:*

$$dp(I|\theta) = \frac{1}{\mathcal{Z}(\theta)} \exp\left[-\frac{1}{2}\ell^2(I)\right] d\mu(I|\theta), \quad (80)$$

where  $\ell(I) = \mathfrak{D}_\theta(I, \bar{I})$  is the separation distance between the points  $\bar{I}$  and  $I$ . Consequently, the knowledge of the metric tensor  $g_{ij}(I|\theta)$  and the point  $\bar{I}$  with maximum information potential  $\mathcal{S}(I|\theta)$  fully determines the distribution function  $dp(I|\theta)$ .

**Proof.** Riemannian gaussian representation (80) is a direct consequence of replacing equation (74) into equation (36). The radius  $\ell$  of the  $n$ -dimensional sphere  $S^n(\bar{I}, \ell)$  referred to in **Theorem 4** and the invariant volume element  $d\mu(I|\theta)$  are purely *geometric notions* derived from the knowledge of the metric tensor  $g_{ij}(I|\theta)$  and the point  $\bar{I}$  with maximum information potential  $\mathcal{S}(I|\theta)$ . Equation (80) evidences that all the statistical description associated with the distribution function (2) can be rephrased in terms of geometric notions derived from the Riemannian structure of the manifold  $\mathcal{M}_\theta$ . ■

**Corollary 6** *For points  $I$  close to the point  $\bar{I}$  with maximum information potential  $\mathcal{S}(I|\theta)$ , the distribution function (36) admits the following gaussian approximation:*

$$dp(I|\theta) \simeq \exp\left[-\frac{1}{2}g_{ij}(\bar{I}|\theta)\Delta I^i\Delta I^j\right] \sqrt{\left|\frac{g_{ij}(\bar{I}|\theta)}{2\pi}\right|} dI. \quad (81)$$

**Proof.** The separation distance  $\ell(I) = \mathfrak{D}_\theta(I, \bar{I})$  can be approximated as follows:

$$\ell^2(I) \simeq g_{ij}(\bar{I}|\theta)\Delta I^i\Delta I^j. \quad (82)$$

This last expression can be directly obtained from definition (40), where  $\Delta I^i = I^i - \bar{I}^i$ . In this approximation level, the normalization condition implies the following estimation for gaussian partition function  $\mathcal{Z}(\theta) \simeq 1$ . ■

## 5. Application examples

### 5.1. Fluctuation geometry of an one-dimensional statistical manifold $\mathcal{M}_\theta$

Let  $dp(I|\theta)$  be a generic parametric family defined on an one-dimensional manifold  $\mathcal{M}_\theta$ . Let us also consider that the admissible values of the stochastic variable in the coordinate representation  $\mathcal{R}_I$  belong to a certain real subset  $(I_{min}, I_{max}) \subset \mathbb{R}$ . Due to its general multidimensional character, the statistical manifold  $\mathcal{P}$  of control parameters  $\theta$  could be a flat or a curved Riemannian manifold. A particular example with a great relevance in statistical and physical applications is the *exponential family*||:

$$dp(I|\theta) = \exp [P(\theta) - \theta^\alpha A_\alpha(I) + B(I)] dI. \quad (83)$$

According to Amari's  $\sigma$ -connections [1]:

$$\Gamma_{\alpha\beta\gamma}^{(\sigma)}(\theta) = \left\langle \left( \frac{\partial v_\beta(\mathcal{I}|\theta)}{\partial \theta^\alpha} + \frac{1-\sigma}{2} v_\alpha(\mathcal{I}|\theta) v_\beta(\mathcal{I}|\theta) \right) v_\gamma(\mathcal{I}|\theta) \right\rangle. \quad (84)$$

the statistical manifold  $\mathcal{P}$  associated with the exponential family (83) is trivially flat when the connection parameter  $\sigma = \pm 1$ . However, the statistical manifold  $\mathcal{P}$  could be a curved manifold for other values of connection parameter  $\sigma$ . An special case is  $\sigma = 0$ , which corresponds to the Levi-Civita connection (43) associated with the inference metric tensor  $g_{\alpha\beta}(\theta) \equiv -\partial^2 P(\theta) / \partial \theta^\alpha \partial \theta^\beta$ . Without mattering about the geometry of the statistical manifold  $\mathcal{P}$ , the one-dimensional statistical manifold  $\mathcal{M}_\theta$  is always *diffeomorphic* to the one-dimensional Euclidean manifold  $\mathbb{E}$ . Clearly, the curvature notion is only admissible for manifolds with dimension  $n \geq 2$ , and hence, the one-dimensional manifold  $\mathcal{M}_\theta$  must exhibit a *flat geometry*. As expected, this type of distribution functions represents the simplest application framework of fluctuation geometry.

The invariant volume element (35) of the one-dimensional manifold  $\mathcal{M}_\theta$  can be rewritten in term of the statistical distance (3) as follows:

$$d\mu(I|\theta) = \sqrt{g_{11}(I|\theta)/2\pi} dI \equiv ds/\sqrt{2\pi}, \quad (85)$$

where  $g_{11}(I|\theta)$  denotes the only component of the metric tensor. One can apply the Riemannian gaussian representation (80) instead of performing the integration of the set of covariant partial differential equations (57). For convenience, let us firstly introduce the coordinate reparametrization  $s(I|\theta) : \mathcal{R}_I \rightarrow \mathcal{R}_s$  defined by the distance  $\mathfrak{D}_\theta(I|\bar{I}) \equiv \ell(I)$  as follows:

$$s(I|\theta) = \begin{cases} -\mathfrak{D}_\theta(I|\bar{I}) & \text{for } I < \bar{I}, \\ \mathfrak{D}_\theta(I|\bar{I}) & \text{for } I \geq \bar{I}, \end{cases} \quad (86)$$

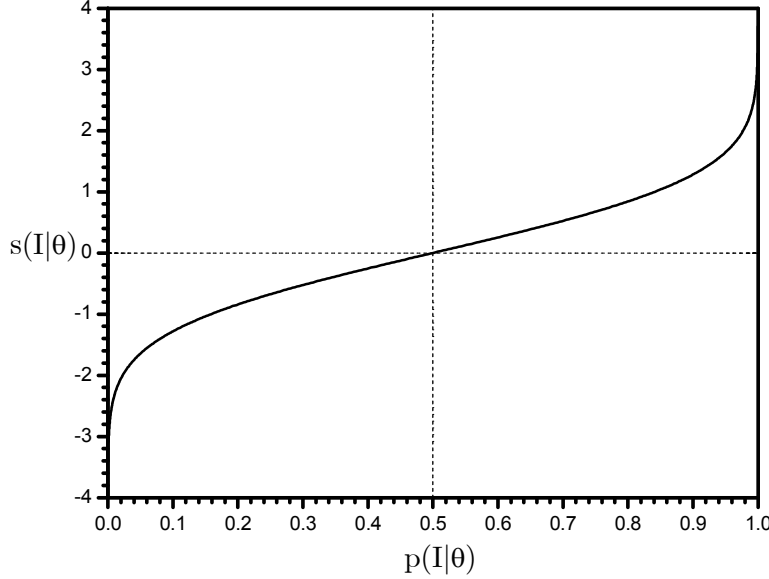
where the metric tensor component  $g_{11}(s|\theta) \equiv 1$ . Using equations (85) and (86), Riemannian gaussian representation (80) can be expressed in the coordinate representation  $\mathcal{R}_s$  as:

$$dp(I|\theta) = \rho(I|\theta) dI \equiv \frac{1}{\mathcal{Z}(\theta)} e^{-\frac{1}{2}s^2} \frac{ds}{\sqrt{2\pi}}. \quad (87)$$

As discussed in Appendix A, the previous expression allows a straightforwardly derivation of the reparametrization function  $s(I|\theta)$ . Introducing the *cumulant distribution function*  $p(I|\theta)$ :

$$p(I|\theta) = \int_{I_{min}}^I dp(I'|\theta) dI', \quad (88)$$

|| Classical statistical ensembles as canonical and Gran canonical ensembles belong to the exponential family (83).



**Figure 1.** Dependence of the reparametrization function  $s(I|\theta)$  versus the cumulant distribution function  $p(I|\theta)$ . Notice that the value corresponding to the maximum information potential  $s(\bar{I}|\theta) = 0$  takes place when the cumulant function  $p(\bar{I}|\theta) = 1/2$ .

the reparametrization function  $s(I|\theta)$  is given by:

$$s(I|\theta) = \Phi^{-1} [p(I|\theta)]. \quad (89)$$

Here,  $\Phi^{-1}(z)$  is the inverse of the function  $\Phi(z)$ :

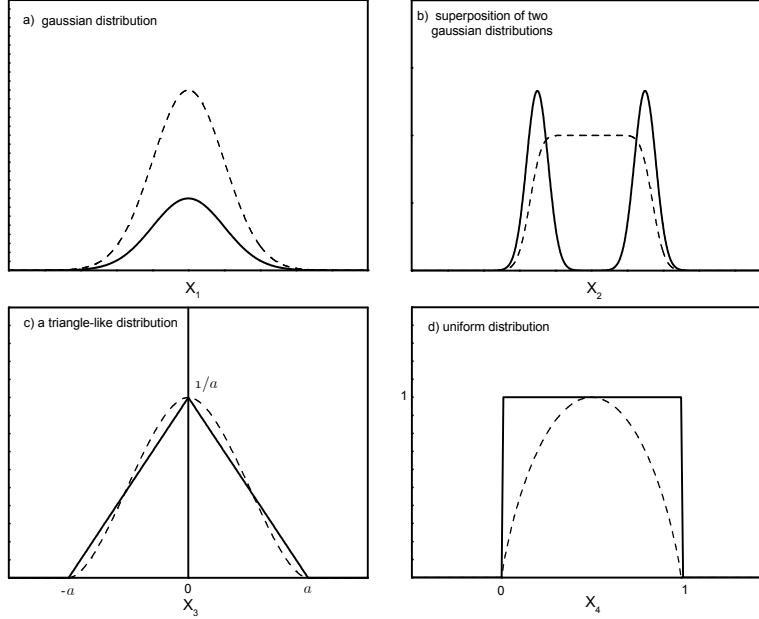
$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}s^2} ds \equiv \frac{1}{2} \left( 1 + \operatorname{erf}(z/\sqrt{2}) \right), \quad (90)$$

with  $\operatorname{erf}(z)$  being the *error function*:

$$\operatorname{erf}(z) = \sqrt{\frac{2}{\pi}} \int_0^z e^{-x^2} dx. \quad (91)$$

The dependence between the reparametrization function  $s(I|\theta)$  and the cumulant distribution function  $p(I|\theta)$  is illustrated in figure 1. By definition of the reparametrization function (86), the point  $\bar{I}$  with maximum information potential  $\mathcal{S}(I|\theta)$  corresponds to the condition  $s(\bar{I}|\theta) = 0$ . According to equation (89), the information potential  $\mathcal{S}(I|\theta)$  exhibits its maximum value at the point  $\bar{I}$  where the cumulant distribution function (88) reaches the value  $p(\bar{I}|\theta) = 1/2$ . Moreover, the admissible values of the variable  $s$  belong to the entire real space  $\mathbb{R}$ ,  $-\infty \leq s \leq +\infty$ . The normalization of the gaussian distribution (87) implies that the gaussian partition function  $\mathcal{Z}(\theta) \equiv 1$ . Thus, the information potential  $\mathcal{S}(I|\theta)$  is given by:

$$\mathcal{S}(I|\theta) \equiv -s^2(I|\theta)/2, \quad (92)$$



**Figure 2.** Comparison between probability density function  $\rho(I|\theta)$  (solid line) and probability weight  $\omega(I|\theta)$  (dashed line) for some simple distribution functions: Panel a) A gaussian distribution defined on the one-dimensional real space  $\mathbb{R}$ . Panel b) The superposition of two different gaussian distributions defined on the one-dimensional real space  $\mathbb{R}$ . Panel c) a triangle-like distribution defined on the real segment  $[-a, a]$ . Panel d) A uniform distribution defined on the real segment  $[0, 1]$ . Despite their different appearance, all these distributions are *diffeomorphic*, that is, they are equivalent from the viewpoint of fluctuation geometry.

which is a non positive function that diverges at the boundary  $\partial\mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$  (at the boundary points  $I_{min}$  and  $I_{max}$  in the coordinate representation  $\mathcal{R}_I$ ). The only component of the metric tensor  $g_{ii}(I|\theta)$  can be expressed as follows:

$$g_{11}(I|\theta) = 2\pi\rho^2(I|\theta) \exp[s^2(I|\theta)]. \quad (93)$$

For illustrative purposes, it is shown in figure 2 a comparison between the probability density function  $\rho(I|\theta)$  and the probability weight  $\omega(I|\theta) = \rho(I|\theta)\sqrt{|2\pi g^{11}(I|\theta)|} \equiv \exp[S(I|\theta)]$  for some simple distribution functions  $dp(I|\theta)$ . While the probability density  $\rho(I|\theta)$  can be a *multimodal function* in certain coordinate representations  $\mathcal{R}_I$  (as the case shown in panel 2.b), the probability weight  $\omega(I|\theta)$  is always a *monomodal scalar function* as a consequence of **Theorem 3**.

Summarizing, the present analysis demonstrates that any parametric family  $dp(I|\theta)$  defined on an one-dimensional statistical manifold  $\mathcal{M}_\theta$  can always be mapped onto the gaussian distribution function (87) using the reparametrization function (89). Consequently, all these distribution functions (as the ones shown in figure 2) are *diffeomorphic* among them. In other words, all them can be regarded as an *abstract gaussian distribution function* defined on the Euclidean manifold  $\mathbb{E}$ , but expressed in different coordinate representations  $\mathcal{R}_I$ .<sup>¶</sup> On the other hand, the relationship between

<sup>¶</sup> The concept of *diffeomorphic distribution functions* will be considered in subsection 6.2 to discuss the notion of *intrinsic differential entropy* of a statistical manifold  $\mathcal{M}_\theta$ .



the reparametrization function  $s(I|\theta)$  and the cumulant distribution function  $p(I|\theta)$  evidences the *purely statistical significance* of the distance notion (3). As expected, fluctuation geometry establishes a direct correspondence between the geometrical description of the statistical manifold  $\mathcal{M}_\theta$  and its probabilistic description. For example, the following geometrical and probabilistic inequalities:

$$\mathfrak{D}_\theta(I|\bar{I}) < \varepsilon \text{ and } |p(I|\theta) - 1/2| \leq \frac{1}{2}\text{erf}(\varepsilon/\sqrt{2}) \quad (94)$$

are fully equivalent.

### 5.2. Generalization to the $n$ -dimensional Euclidean manifold $\mathbb{E}^n$

Let us suppose that the parametric family  $dp(I|\theta)$  can be factorized into independent distribution functions  $dp^{(i)}(I^i|\theta)$  for each stochastic variable  $I^i$ :

$$dp(I|\theta) = \prod_i dp^{(i)}(I^i|\theta). \quad (95)$$

Accordingly, its associated  $n$ -dimensional statistical manifold  $\mathcal{M}_\theta$  can be decomposed as the *external product* of the set of one-dimensional statistical manifolds  $\{\mathcal{E}_\theta^i\}$  as follows:

$$\mathcal{M}_\theta = \mathcal{E}_\theta^1 \otimes \mathcal{E}_\theta^2 \dots \otimes \mathcal{E}_\theta^n, \quad (96)$$

and hence,  $\mathcal{M}_\theta$  is diffeomorphic to the  $n$ -dimensional Euclidean manifold  $\mathbb{E}^n$ . The results obtained in the previous subsection are straightforwardly extended to the present situation considering that the information potential  $\mathcal{S}(I|\theta)$  is additive:

$$\mathcal{S}(I|\theta) = \sum_{i=1}^n \mathcal{S}^{(i)}(I^i|\theta), \quad (97)$$

while the metric tensor  $g_{ij}(I|\theta)$  is diagonal:

$$ds^2 = \sum_{i=1}^n g_{ii}(I^i|\theta)(dI^i)^2. \quad (98)$$

Here, the functions  $\mathcal{S}^{(i)}(I^i|\theta)$  and  $g_{ii}(I^i|\theta)$  are obtained from the probability densities  $\rho^{(i)}(I^i|\theta)$  using equations (88), (89), (92) and (93).

## 6. Implications of fluctuation geometry in statistics and physics

### 6.1. Comparison between fluctuation geometry and inference geometry

As naturally expected, the distance notion of inference geometry (1) allows to define a *statistical distance*  $\mathfrak{D}(\vartheta|\theta)$  *between two members* (two different distribution functions) of the parametric family (2), e.g.: considering the arc-length of the geodesics that connects the points  $\theta$  and  $\vartheta \in \mathcal{P}$ . According to asymptotic formula (31), this statistical distance is associated with the *distinguishing probability* of these distribution functions during a statistical inferential procedure. Conversely, the distance notion of fluctuation geometry (3) allows to define a *statistical distance*  $\mathfrak{D}_\theta(I_1|I_2)$  *between two sets of values of the stochastic variables  $I$ , which are described by a given member* (a specific distribution function) *of the parametric family* (2). At first glance, the approximation formula (81) can be regarded as a *counterpart expression* of the asymptotic distribution (31) of inference geometry.

Such an analogy clarifies the *relevance* of the distance notion (3): this second statistical distance is associated with the *occurrence probability* of a small fluctuation  $\Delta I = I - \bar{I}$  around the point  $\bar{I}$  with maximum information potential  $\mathcal{S}(I|\theta)$ . Remarkably, the asymptotic formula (81) is the crudest approximation of the Riemannian gaussian representation (80). According to this rigorous result of fluctuation geometry, the statistical distance  $\mathfrak{D}_\theta(I|\bar{I})$  is simply a measure of the *relative occurrence probability* in regard to the point with maximum information potential  $\bar{I}$ :

$$\mathfrak{D}_\theta^2(I|\bar{I}) \equiv -2 \log \left[ \frac{\omega(I|\theta)}{\omega(\bar{I}|\theta)} \right], \quad (99)$$

with  $\omega(I|\theta)$  being the probability weight (48).

Although inference geometry and fluctuation geometry are two counterpart approaches, they provide *different qualitative information* about the statistical properties of a given parametric family (2). On one hand, the first theory provides geometric information concerning to the inference of the control parameters  $\theta$  of a given parametric family (2). On the other hand, the second theory provides geometric information about the fluctuating behavior of the stochastic variables  $I$  for a given distribution function of the parametric family (2). Noteworthy that the term *geometric information* has a special meaning here: these geometric theories consider those properties of the statistical manifolds  $\mathcal{P}$  and  $\mathcal{M}_\theta$  that are independent on their specific coordinate representations  $\mathcal{R}_\theta$  and  $\mathcal{R}_I$ . Additionally, these two statistical geometries differ in regard to their *application frameworks*. Inference geometry only demands the *continuous character* of the statistical manifold  $\mathcal{P}$  of control parameters  $\theta$ . Therefore, this type of geometry can be introduced for a *parametric family of distribution functions*  $p(X|\theta)$  defined on a set of discrete variables  $X = \{X_k\}$ :

$$p(X|\theta) = \{p(X_k|\theta) | k \in \mathbb{Z}\}. \quad (100)$$

Conversely, fluctuation geometry only demands the *continuous character* of the manifold  $\mathcal{M}_\theta$  of stochastic variables  $I$ . Therefore, this geometry can be introduced for a *continuous distribution function without control parameters*:

$$dp(I) = \rho(I)dI. \quad (101)$$

As expected, the simultaneous definition of inference geometry and fluctuation geometry is only possible for parametric families (2) defined on continuous statistical manifolds  $\mathcal{M}_\theta$  and  $\mathcal{P}$ .

A simple look to equations (28) and (57) allows us to realize that these geometric theories have a different *amenable character*. In particular, the metric tensor  $g_{\alpha\beta}(\theta)$  of inference geometry (28) is very easy to obtain, either from the analytical or numerical calculation of these integrals. Conversely, the metric tensor  $g_{ij}(I|\theta)$  of fluctuation geometry should be obtained solving a set of covariant partial differential equations (57), whose admissible solutions must obey certain boundary conditions. Actually, this latter mathematical procedure can be a hard task for a manifold  $\mathcal{M}_\theta$  with a nontrivial geometry. However, once obtained the metric tensors  $g_{\alpha\beta}(\theta)$  and  $g_{ij}(I|\theta)$ , the amenable character of these geometric theories changes in a radical way: *fluctuation geometry turns much more amenable than inference geometry*. For example, a modest mathematical effort has been devoted to arrive at the rigorous **Theorems 2-4** and their associated implications as the Riemannian gaussian representation (80). Conversely, inference geometry has to deal with a *very serious difficulty*: there not exist a general way to relate a given parametric family (2) and the *unbiased estimators*  $\hat{\theta}$  of its control parameters  $\theta$ . Even the calculation of the distribution function (30) of the *efficient unbiased estimators*  $\hat{\theta}_{eff}$  is not a easy task. In particular, the asymptotic distribution (31) is a direct application of the *central limit theorem*, that is, an approximation formula for a statistical inference with a large but finite number  $m$  of outcomes

$\mathcal{I} = \{I^{(1)}, I^{(2)}, \dots, I^{(m)}\}$ . A very important question during the historical development of inference geometry is the so-called *higher-order asymptotic theory of statistical estimation* [1], where a fundamental task is the improvement of approximation formula (31) considering a  $1/m$ -power expansion based on the intrinsic geometry of the manifold  $\mathcal{P}$ . A counterpart of Riemannian gaussian representation (80) in the framework of inference geometry is unknown in the literature, at least, from the knowledge of the present author.

## 6.2. On the notion of information entropy for a continuous distribution

As discussed elsewhere [21], the *information entropy*  $S[p|\theta]$  associated with the discrete distribution function (100) is written as follows:

$$S[p|\theta] = - \sum_k p(X_k|\theta) \log p(X_k|\theta). \quad (102)$$

Conceptually, information entropy is considered as a measure of the *unpredictability* associated with a random variable  $X$ . Interestingly, its counterpart extension for a continuous distribution function:

$$dQ(I) = q(I)dI, \quad (103)$$

the so-called *differential entropy*:

$$S_{de}^{(\mathcal{R}_I)}[Q|\mathcal{M}_\theta] = - \int_{\mathcal{M}_\theta} q(I) \log q(I) dI, \quad (104)$$

undergoes an important geometric inconsistency: its definition crucially depends on the coordinate representation  $\mathcal{R}_I$  of the statistical manifold  $\mathcal{M}_\theta$ :

$$S_{de}^{(\mathcal{R}_I)}[Q|\mathcal{M}_\theta] \neq S_{de}^{(\mathcal{R}_\Theta)}[Q|\mathcal{M}_\theta] = - \int_{\mathcal{M}_\theta} q(\Theta) \log q(\Theta) d\Theta. \quad (105)$$

In general, the expectation values of scalar functions defined on the statistical manifold  $\mathcal{M}_\theta$  are only independent on the coordinate representations:

$$A(I) = A(\Theta) \Rightarrow \langle A(I) \rangle = \langle A(\Theta) \rangle. \quad (106)$$

However, the probability density function  $q(I)$  is simply a *tensorial density*, whose values and general mathematical behavior crucially depend on the concrete coordinate representation  $\mathcal{R}_I$  of the statistical manifold  $\mathcal{M}_\theta$ . Consequently, the consideration of the quantity  $\mathcal{I}_c(I) = -\log q(I)$  as a local measure of the *information content* is ill-defined from the geometric viewpoint because of it violates the requirement of covariance under the coordinate reparametrization  $\Theta(I) : \mathcal{R}_I \rightarrow \mathcal{R}_\Theta$  of the statistical manifold  $\mathcal{M}_\theta$ . Despite their apparent similarity, the differential geometry (104) is not a good generalization of the statistical entropy (102) for the framework of continuous distribution functions. For example, the differential entropy (104) does not obey other properties of its discrete counterpart (102), in particular, the positive definition  $S[p|\theta] \geq 0$ .

An attempt to overcome some of the above inconsistencies was developed by Jaynes [22]. According to this author, the correct formula for the information entropy of a continuous distribution function can be derived taking the *limit of increasingly dense discrete distributions*. Specifically, Jaynes proposed to start from a set of  $n$  discrete points  $\mathcal{S}_n = \{I_i\} \subset \mathcal{M}_\theta$ , which density  $\gamma_n(I) = n^{-1} \sum_{i=1}^n \delta(I - I_i)$  approaches a certain function  $\gamma(I)$  in the limit  $n \rightarrow \infty$ :

$$\gamma(I) = \lim_{n \rightarrow \infty} \gamma_n(I). \quad (107)$$

The density function  $\gamma(I)$  is referred to as the *invariant measure*. Combining the previous argument and the discrete definition of information entropy (102), this author arrived at the following correction for the differential entropy:

$$\mathcal{S}_{de}^\gamma [Q|\mathcal{M}_\theta] = - \int_{\mathcal{M}_\theta} q(I) \log \left[ \frac{q(I)}{\gamma(I)} \right] dI. \quad (108)$$

At first glance, the present formula is similar to but conceptually distinct from the (negative of the) *Kullback-Leibler divergence* [20]:

$$D_{KL}(Q|P) = \int_{\mathcal{M}_\theta} q(I) \log \left[ \frac{q(I)}{p(I)} \right] dI. \quad (109)$$

As many other divergences considered in statistics [20], Kullback-Leibler divergence (109) is a measure of the *separation* of a distribution function  $dQ(I) = q(I)dI$  to a reference distribution  $dP(I) = p(I)dI$ . In the formula (108), however, the invariant measure  $\gamma(I)$  need not to be a *probability density*, but simply a density. In particular, it need not satisfied the normalization condition:

$$\int_{\mathcal{M}_\theta} \gamma(I) dI \neq 1. \quad (110)$$

Although Jaynes' differential entropy (108) is invariant under coordinate reparametrizations, the success achieved with this correction formula is only partial. In fact, Jaynes was unable to provide a general criterium to precise the invariant measure  $\gamma(I)$  for a concrete application. Referring to this ambiguity, he recognized that ... *the following arguments can be made as rigorous as we please, but at considerable sacrifice of clarity* [22]. Remarkably, the pre-existence of a *Riemannian structure* defined on the statistical manifold  $\mathcal{M}_\theta$  introduces a natural choice for the invariant measure  $\gamma(I)$ . While the probability density  $q(I)$  of a distribution function  $dQ(I)$  depends on the coordinate representation  $\mathcal{R}_I$ , the notion of *probability weight*<sup>+</sup>:

$$q_g(I) = q(I) \sqrt{|2\pi g^{ij}(I|\theta)|} \quad (111)$$

represents a scalar function defined on the statistical manifold  $\mathcal{M}_\theta$ . Using the probability weight  $q_g(I)$  instead of the probability density  $q(I)$ , the quantity  $\mathfrak{I}_c(I|\theta) = -\log q_g(I)$  can be introduced as a *local invariant measure of the information content*. Thus, a more appropriate generalization of information entropy for a continuous distribution function is given by:

$$\mathcal{S}_{de}^g [Q|\mathcal{M}_\theta] = \langle \mathfrak{I}_c(I|\theta) \rangle = - \int_{\mathcal{M}_\theta} q_g(I) \log q_g(I) d\mu(I|\theta), \quad (112)$$

where the index  $g$  denotes the Riemannian structure of the statistical manifold  $\mathcal{M}_\theta$ . Noteworthy that equation (112) is a particular case of Jaynes' differential entropy (108), where the invariant measure  $\gamma(I)$  is determined by the metric tensor  $g_{ij}(I|\theta)$  of the statistical manifold  $\mathcal{M}_\theta$ :

$$\gamma(I) = \sqrt{|g_{ij}(I|\theta)/2\pi|} \equiv 1/\sqrt{|2\pi g^{ij}(I|\theta)|}. \quad (113)$$

According to Jaynes' argument, the invariant measure (113) can be obtained as the limit of increasingly dense subset of points  $\mathcal{S}_n$  that are *uniformly distributed* on the statistical manifold  $\mathcal{M}_\theta$ , that is, a distribution function whose probability weight  $\gamma_g(I|\theta) \equiv 1$ .

The geometric differential entropy (112) depends both on the distribution function  $dQ(I)$  as well as the Riemannian structure of the statistical manifold  $\mathcal{M}_\theta$ . According to postulates of

<sup>+</sup> The notion of probability weight was considered in equation (48) to introduce the information potential  $\mathcal{S}(I|\theta)$ .

fluctuation geometry, the Riemannian structure of the statistical manifold  $\mathcal{M}_\theta$  is associated with a *reference distribution function*  $dp(I|\theta)$ , specifically, the distribution function (36) derived from the knowledge of the information potential  $\mathcal{S}(I|\theta)$ . Therefore, it is worth to distinguish between two different notions of differential entropy:

- The differential entropy  $\mathcal{S}_{de}^g[Q|\mathcal{M}_\theta]$  of an arbitrary distribution function  $dQ(I)$  defined on a statistical manifold  $\mathcal{M}_\theta$ , which is endowed of a *pre-existent* Riemannian structure.
- The notion of *intrinsic differential entropy*  $\mathcal{S}_{de}^g[\mathcal{M}_\theta]$  of a statistical manifold  $\mathcal{M}_\theta$ , that is, the differential entropy  $\mathcal{S}_{de}^g[p|\mathcal{M}_\theta]$  of the distribution function  $dp(I|\theta)$  associated with the Riemannian structure of the statistical manifold  $\mathcal{M}_\theta$ .

The intrinsic differential entropy  $\mathcal{S}_{de}^g[\mathcal{M}_\theta]$  of the statistical manifold  $\mathcal{M}_\theta$  is given by the negative of the expectation value of the information potential  $\mathcal{S}(I|\theta)$ :

$$\mathcal{S}_{de}^g[\mathcal{M}_\theta] \equiv -\langle \mathcal{S}(I|\theta) \rangle, \quad (114)$$

which can be rewritten using equation(74) as follows:

$$\mathcal{S}_{de}^g[\mathcal{M}_\theta] = \frac{1}{2} \langle \mathfrak{D}_\theta^2(I|\bar{I}) \rangle - \mathcal{P}(\theta). \quad (115)$$

Accordingly,  $\mathcal{S}_{de}^g[\mathcal{M}_\theta]$  is a *global geometric measure* of an statistical manifold  $\mathcal{M}_\theta$ , which depends on its topological properties and Riemannian structure, as well as the position of the point  $\bar{I}$  with maximum information potential. In particular, if the statistical manifold  $\mathcal{M}_\theta$  can be decomposed into two independent Riemannian manifolds  $\mathcal{A}$  and  $\mathcal{B}$  as  $\mathcal{M}_\theta = \mathcal{A} \otimes \mathcal{B}$ , its intrinsic differential entropy  $\mathcal{S}_{de}^g[\mathcal{M}_\theta]$  is additive:

$$\mathcal{S}_{de}^g[\mathcal{M}_\theta] = \mathcal{S}_{de}^{g_A}[\mathcal{A}] + \mathcal{S}_{de}^{g_B}[\mathcal{B}], \quad (116)$$

where  $g_A$  and  $g_B$  denote their respective Riemannian structures.

Before we end this section, it is worth remarking that the requirement of covariance is a strong constraint. The existence of this symmetry in the differential entropy (112) implies that *diffeomorphic distribution functions exhibit the same value of their intrinsic differential entropies*. As already commented, all distribution functions illustrated in figure 2 are diffeomorphic. In fact, their statistical manifolds  $\mathcal{M}_\theta$  are diffeomorphic to the one-dimensional Euclidean manifold  $\mathbb{E}$ . Thus, their respective intrinsic differential entropies exhibit the same value  $\mathcal{S}_{de}^g[\mathbb{E}] = 1/2$ . This result is easy to obtain using the gaussian distribution (87) associated with the coordinate representation  $\mathcal{R}_s$  of the statistical manifold  $\mathcal{M}_\theta$ . Moreover, the  $n$ -dimensional manifold  $\mathcal{M}_\theta$  associated with a distribution function obeying the decomposition (95) has an intrinsic information entropy  $\mathcal{S}_{de}^g[\mathbb{E}^n] = n/2$ , which is a direct consequence of the property (116)

### 6.3. Riemannian reformulation of Einstein fluctuation theory

Classical fluctuation theory starts from *Einstein postulate* [23]:

$$dp(I|\theta) = \mathcal{A}e^{S(I|\theta)}dI, \quad (117)$$

which describes the fluctuating behavior of a set of macroscopic observables  $I$  in an equilibrium situation driven by certain control parameters  $\theta^*$ . Here,  $S(I|\theta)$  denotes the entropy of a closed system, while  $\mathcal{A}$  is a normalization constant. Since the function  $\rho(I|\theta) = \mathcal{A}e^{S(I|\theta)}$  is a tensorial

\* The Boltzmann's constant  $k$  is set as the unity,  $k = 1$ .

density, the entropy  $S(I|\theta)$  considered in Einstein postulate (117) behaves under a coordinate reparametrization  $\Theta(I) : \mathcal{R}_I \rightarrow \mathcal{R}_\Theta$  as follows:

$$S(\Theta|\theta) = S(I|\theta) - \log \left| \frac{\partial \Theta}{\partial I} \right|, \quad (118)$$

and hence, it is not a scalar function. This feature is a direct contradiction with the thermodynamic relevance of entropy as a *state function*<sup>‡</sup>. Consequently, Einstein postulate (117) is *ill-defined from the geometric viewpoint*.

As already discussed in a recent paper [18], the previous inconsistency disappears when one redefines Einstein postulate into the covariant form (36). Thus, the entropy of a closed system should be identified with the information potential  $\mathcal{S}(I|\theta)$  of fluctuation geometry up to the precision of an additive constant. As expected, the covariant distribution function (36) also depends on the Riemannian structure of the statistical manifold  $\mathcal{M}_\theta$  of the macroscopic observables  $I$ . Such an ambiguity disappears considering the relationship between the metric tensor  $g_{ij}(I|\theta)$  and the negative of the *covariant Hessian* of the scalar entropy  $\mathcal{S}(I|\theta)$ , equation (51). Consequently, the system fluctuating behavior is fully determined by the knowledge of the entropy  $\mathcal{S}(I|\theta)$ . From the physical viewpoint, the constraint (51) between the metric tensor  $g_{ij}(I|\theta)$  and the entropy  $\mathcal{S}(I|\theta)$  is very relevant. In fact, such a choice ensures the *geodesic character* of the *system hydrodynamic equations* [18]:

$$\frac{dI^i(s)}{ds} = -v^i(I|\theta). \quad (119)$$

Here, the parameter  $s$  denotes the arc-length of the curve of hydrodynamic relaxation  $I(s) \in \mathcal{M}_\theta$ . This curve can be characterized by the tangent vector field  $\xi(s)$  with contravariant components:

$$\xi^i(s) = \frac{dI^i(s)}{ds}, \quad (120)$$

which is unitary vector field,  $g_{ij}(I|\theta)\xi^i(s)\xi^j(s) \equiv 1$ . The tangent vector field  $\xi(s)$  is oriented in the same direction of the *generalized restituting force*  $\zeta(I|\theta)$  with covariant components  $\zeta_i(I|\theta) = \partial\mathcal{S}(I|\theta)/\partial I^i \equiv -\psi_i(I|\theta)$ . As expected, the generalized restituting force is directed towards the equilibrium configuration  $\bar{I}$  (the point with maximum entropy), whose direction is determined by the negative of the unitary vector field  $v^i(I|\theta)$  introduced in equation (75). As already shown in the proof of **Theorem 4**, the unitary vector field  $v^i(I|\theta)$  obeys geodesic differential equations (52). In physics, the geodesics are regarded as the *natural motions* in theories with a Riemannian formulation, as example, in general relativity theory. This nontrivial result establishes an interesting connection of the present Riemannian reformulation of equilibrium fluctuation theory and its possible nonequilibrium generalization.

#### 6.4. Relation with Ruppeiner's geometry of thermodynamics

Ruppeiner proposed in the past a Riemannian geometry for thermodynamics [24]. Despite its large history in the literature, Ruppeiner's geometry undergoes some inconsistencies. The most important is that this approach assumes the scalar character of entropy  $S(I|\theta)$  employed in Einstein postulate (117). Ruppeiner recognized himself the restricted applicability of this consideration

<sup>‡</sup> A state function is a property of a system that depends only on its current state, not on the way in which the system acquired that state. Geometrically, the value of a state function should not depend on the coordinate representation employed to describe that state, that is, it should be a *scalar function*.

(see in subsection II.B in his paper in Review of Modern Physics [24]). However, he justified its application as a *good approximation* in the framework of large thermodynamic systems, whose fluctuating behavior can be described within the *gaussian approximation*:

$$dp(I|\bar{I}) \simeq \exp \left[ -\frac{1}{2} g_{ij}(\bar{I}) \Delta I^i \Delta I^j \right] \sqrt{\left| \frac{g_{ij}(\bar{I})}{2\pi} \right|} dI. \quad (121)$$

Here,  $\Delta I^i = I^i - \bar{I}^i$  denotes a small deviation from the most likely (equilibrium) state  $\bar{I}$ , while  $g_{ij}(\bar{I})$  represents the *thermodynamic metric tensor* [24]:

$$g_{ij}(\bar{I}) = -\frac{\partial^2 S(\bar{I}|\theta)}{\partial I^i \partial I^j}. \quad (122)$$

Notice that the thermodynamic metric tensor is expressed here in terms of the most likely state  $\bar{I}$  instead of the control parameters  $\theta$ . Apparently, the goal of this consideration is to justify the relevance of the distance notion:

$$ds_R^2 = g_{ij}(\bar{I}) d\bar{I}^i d\bar{I}^j \quad (123)$$

as a *thermodynamic distance between equilibrium states*. The previous consideration, however, has a restricted applicability even in the framework of large thermodynamic systems. Since the function  $\rho(I|\theta) = \mathcal{A}e^{S(I|\theta)/k}$  is a tensorial density, the probability distribution (117) can exhibit two or more maxima  $\bar{I}$  for certain values of control parameters  $\theta$ . Consequently, it is not possible to guarantee the bijective correspondence between the control parameters  $\theta$  and the equilibrium states  $\bar{I}$  (the most likely values of macroscopic observables). This type of situations is associated with the *phenomenon of ensemble inequivalence* in statistical mechanics, which is observed during the occurrence of *discontinuous phase transitions* [23].

Fluctuation geometry successfully overcome the previous limitations of Ruppeiner's geometry [18]. As already commented, the key considerations are (i) to assume the covariant redefinition of Einstein postulate (36) to ensure the scalar character of the entropy  $\mathcal{S}(I|\theta)$ ; and (ii) to generalize the thermodynamic metric tensor (122) considering the scalar entropy  $\mathcal{S}(I|\theta)$  and the covariant differentiation  $D_i$ :

$$g_{ij}(\bar{I}) = -\frac{\partial^2 S(\bar{I}|\theta)}{\partial I^i \partial I^j} \rightarrow g_{ij}(I|\theta) = -D_i D_j \mathcal{S}(I|\theta). \quad (124)$$

From this viewpoint, Riemannian reformulation of Einstein fluctuation theory arises as a formal improvement of Ruppeiner's geometry. Rewriting the previous relation as follows:

$$g_{ij}(I|\theta) = -\frac{\partial^2 \mathcal{S}(I|\theta)}{\partial I^i \partial I^j} + \Gamma_{ij}^k(I|\theta) \frac{\partial \mathcal{S}(I|\theta)}{\partial I^k}, \quad (125)$$

it is easy to check that the metric tensor  $g_{ij}(I|\theta)$  looks like the thermodynamic metric tensor (122) at the point  $\bar{I}$  with maximum entropy, where  $\partial \mathcal{S}(\bar{I}|\theta)/\partial I^i = 0$ . However, the existence and uniqueness of the point  $\bar{I}$  is now guaranteed by **Theorem 3**. Moreover, the metric tensor  $g_{ij}(I|\theta)$  of fluctuation geometry is well-defined for any macroscopic state  $I \in \mathcal{M}_\theta$ . While gaussian approximation (121) is only applicable to large thermodynamic systems with a small fluctuating behavior, Riemannian gaussian representation (80) is a rigorous result. Noteworthy that the application of fluctuation geometry in classical fluctuation theory cannot be regarded as a Riemannian approach of thermodynamics, but a *Riemannian approach of classical statistical mechanics*<sup>††</sup>.

<sup>††</sup>Thermodynamics is a macroscopic physical theory that disregards the incidence of fluctuations.

## 7. Final remarks and open problems

Fluctuation geometry was proposed in this work as a counterpart approach of inference geometry. This new geometry allows the introduction of the distance notion  $ds^2 = g_{ij}(I|\theta)dI^i dI^j$  to characterize a *statistical distance* between two different values of the stochastic variables  $I$ , whose behavior is described by a member of a parametric family of continuous distribution functions  $dp(I|\theta) = \rho(I|\theta)dI$ . The metric tensor  $g_{ij}(I|\theta)$  has been derived starting from a set of axioms, which lead to a set of covariant differential equations (57) written in terms of the probability density  $\rho(I|\theta)$ . The main consequence is the possibility to rephrase the probability description in terms of purely geometric notions, as the case of Riemannian gaussian distribution (80). Thus, the statistical description can be equivalently performed using the language of Riemannian geometry, and hence, fluctuation geometry represents an alternative framework for applying the powerful tools of differential geometry for the statistical analysis. As already evidenced, the present approach leads to a reconsideration of the notion of information entropy for a continuous distribution as well as the Riemannian reformulation of Einstein fluctuation theory.

Before we end this section, let us comment some open problems that deserve a special attention in future works. Firstly, it is important to clarify the *existence and uniqueness* of the solution of the covariant differential equations (57). As already evidenced, postulates of fluctuation geometry allow a univocal determination of the Riemannian structure of the statistical manifold  $\mathcal{M}_\theta$  for the application examples discussed in this work, which exhibit a trivial Euclidean (flat) geometry. Thus, it is necessary to check if such an existence and uniqueness are preserved in a statistical manifold  $\mathcal{M}_\theta$  with a more complex Riemannian structure.

Secondly, one expects that the *curvature notion* of manifold  $\mathcal{M}_\theta$  should play a fundamental role from the statistical viewpoint. Some basic arguments suggest that curvature should be associated with the notion of *statistical correlations*. Both curvature notion and the statistical correlations, as example, can only be defined when the dimension  $n$  of the statistical manifold  $\mathcal{M}_\theta$  is equal or larger than two. Interestingly, the existence of a decomposition (95) in a parametric family implies the flat character of the statistical manifold  $\mathcal{M}_\theta$ . This possible relevance of the curvature notion of the statistical manifold  $\mathcal{M}_\theta$  is consistent with some physical analogies. General Relativity theory, as example, identifies gravitational interaction with the curvature of the space-time  $\mathbb{M}^4$ . In the framework of statistical theories as quantum mechanics, the statistical correlations can be regarded as the counterpart of interactions. The gas of non-interacting particles obeying Fermi-Dirac statistics, in particular, manifests *effective repulsion forces* as consequence of the *inter-particle correlations* associated with Pauli's exclusion principle. By analogy, a non-vanishing curvature of a statistical manifold as  $\mathcal{M}_\theta$  would be associated with the existence of *irreducible statistical correlations*. The analysis of this conjecture will be the main interest of a forthcoming paper.

A third question is to analyze how deep is the analogy between inference geometry and fluctuation geometry. Specifically, it is natural to wonder if each result obtained in one of these theories has a counterpart relation in the other theory. Gaussian approximation (81), as example, can be regarded as a counterpart result of the asymptotic distribution (31) of inference theory. Starting from the fact that gaussian distribution (81) admits the exact improvement (80), the underlying analogy strongly suggests the following improvement:

$$dQ^m(\vartheta|\theta) = \frac{1}{\mathcal{Z}(\theta)} \exp \left[ -\frac{1}{2} \ell^2(\vartheta) \right] \sqrt{\frac{g_{\alpha\beta}(\vartheta)}{2\pi}} d\vartheta \quad (126)$$

for the asymptotic distribution (31) of inference theory. Here,  $\ell(\vartheta)$  should represent the distance  $\mathfrak{D}(\vartheta, \theta)$  between the points  $\vartheta$  and  $\theta$  calculated with the metric tensor  $g_{\alpha\beta}(\theta)$  defined on the statistical



manifold  $\mathcal{P}$  of control parameters  $\theta$ . It would be interesting to analyze this conjecture. Finally, it would be interesting to analyze other implications of fluctuation geometry in the framework of information theory [21]. A question with a special interest is the application of the notion of intrinsic differential entropy to the problem of *maximum information distributions* [25].

### Appendix A. Derivation of the reparametrization function $s(I|\theta)$

Let us denote by  $s_1$  and  $s_2$  the coordinates that correspond to the boundary points  $I_{min}$  and  $I_{max}$  in the new coordinate representation  $\mathcal{R}_s$ . The integration of equation (87) yields the following relation:

$$p(I|\theta) = \frac{\Phi(s) - \Phi(s_1)}{\Phi(s_2) - \Phi(s_1)}. \quad (\text{A.1})$$

Here,  $p(I|\theta)$  represents the cumulant distribution function (88) and  $\Phi(z)$  the function (90). Moreover, it was taken into account that the normalization condition of the distribution function  $dp(I|\theta)$  implies the relation  $\mathcal{Z}(\theta) \equiv \Phi(s_2) - \Phi(s_1)$ . Introducing the inverse function  $\Phi^{-1}(z)$ , the reparametrization function  $s(I|\theta) : \mathcal{R}_I \rightarrow \mathcal{R}_s$  can be expressed as follows:

$$s(I|\theta) = \Phi^{-1}[\phi + \sigma p(I|\theta)], \quad (\text{A.2})$$

where  $\phi = \Phi(s_1)$  and  $\sigma = \Phi(s_2) - \Phi(s_1)$ . Considering the expression (74), the information potential  $\mathcal{S}(I|\theta)$  can be expressed as follows:

$$\mathcal{S}(I|\theta) = -\log \sigma - \frac{1}{2} s^2(I|\theta). \quad (\text{A.3})$$

Both the reparametrization function (A.2) and the information potential (A.3) depend on the nonnegative constants  $\phi$  and  $\sigma$ , whose values are determined by the boundary points  $s_1$  and  $s_2$  in the coordinate representation  $\mathcal{R}_s$ . However, a careful analysis reveals that these parameters cannot admit arbitrary values. Firstly, one should notice that the information potential (A.3) is everywhere finite when  $|s_{1,2}| < \infty$ . Taking into account the relationship between the information potential  $\mathcal{S}(I|\theta)$  and the probability density  $\rho(I|\theta)$ :

$$\rho(I|\theta) \equiv \exp[\mathcal{S}(I|\theta)] \sqrt{\left| \frac{g_{11}(I|\theta)}{2\pi} \right|}, \quad (\text{A.4})$$

the vanishing of the probability density  $\rho(I|\theta)$  at the boundary  $\partial\mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$  (**Axiom 4**) implies the vanishing of the metric tensor  $g_{11}(I|\theta)$  on the boundary  $\partial\mathcal{M}_\theta$ . However, the metric tensor  $g_{11}(I|\theta)$  should be non-vanishing everywhere, even, on the boundary  $\partial\mathcal{M}_\theta$  of the statistical manifold  $\mathcal{M}_\theta$ . The existence of the contravariant metric tensor  $g^{ij}(I|\theta)$ , in particular, demands the non-vanishing of the metric tensor determinant  $|g_{ij}(I|\theta)|$ . For an one dimensional manifold  $\mathcal{M}_\theta$ , this last requirement implies  $0 < |g_{11}(I|\theta)| < \infty$ . The only way to fulfil such a requirement is to impose the constraints  $\phi = 0$  and  $\sigma = 1 \Leftrightarrow s_1 = -\infty$  and  $s_2 = +\infty$ , which leads to equation (89).

### Acknowledgements

Velazquez thanks the financial support of CONICYT/Programa Bicentenario de Ciencia y Tecnología PSD 65 (Chilean agency).

## References

- [1] Amari Sh 1990 *Geometrical Methods in Statistics: 22 Lecture notes in statistics* (Springer Verlag).
- [2] Fisher R A 1922 *Phil. Trans. R. Soc.* **222** 309-368.
- [3] Rao C R 1945 *Bull. Calcutta Math. Soc.* **37** 81-91.
- [4] Frieden B R 1998 *Physics from Fisher Information: A Unification* (Cambridge University Press).
- [5] Brody D C and Hughston L P 1996 *Phys. Rev. Lett.* **7** 2851.
- [6] Barndorff-Nielsen O E and Gill R D 2000 *J. Phys. A* **30** 4481-4490.
- [7] Gibilisco P and Isola T 2006 *Ann. Inst. Stat. Math.* **59** 147159.
- [8] Gibilisco P, Imparato D and Isola T 2007 *J. Math. Phys.* **48** 072109.
- [9] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam).
- [10] Caianiello E R 1986 in *Frontiers of Non-Equilibrium Statistical Physics* ed G T Moore and M O Scully (Plenum, New York) pp. 453-464.
- [11] Braunstein S L 1993 in *Symposium on the Foundations of Modern Physics* Busch P, Lahti P, and Mittelstaedt P (eds.) (World Scientific, Singapore) pp. 106.
- [12] Janke W *et al* 2004 *Physica A* **336** 181-186.
- [13] Brody D C and Hook D W 2009 *J. Phys. A: Math. Theor.* **42** 023001.
- [14] Crooks G E 2007 *Phys. Rev. Lett.* **99** 100602.
- [15] Velazquez L and Curilef S 2009 *J. Phys. A: Math. Theor.* **42** 095006.
- [16] Velazquez L and Curilef S 2009 *Mod. Phys. Lett. B* **23** 3551.
- [17] Velazquez L and Curilef S 2010 *J. Stat. Mech.* **P12031**.
- [18] Velazquez L 2011 *J. Stat. Mech.* **P11007**.
- [19] Berger A 2002 *A panoramic view of Riemannian geometry* (Springer).
- [20] Pardo L 2006 *Statistical inference based on divergence measures* (Taylor & Francis Group, LLC).
- [21] Thomas M. and Cover Joy A 2006 *Elements of Information Theory, Wiley Interscience* (John Wiley & Sons, Inc., Hoboken, New Jersey).
- [22] Jaynes E T 1963, *Information Theory and Statistical Mechanics*, in *Statistical Physics*, K. Ford (ed) (Benjamin, New York) p. 181.
- [23] Reichl L E 1980 *A modern course in Statistical Mechanics*, (Univ. Texas Press, Austin).
- [24] Ruppeiner G 1979 *Phys. Rev. A* **20** 1608; 1995 *Rev. Mod. Phys.* **67** 605.
- [25] Uffink J 1995 *Studies in History and Philosophy of Modern Physics* **26B** 223-261; 1996 *Studies in History and Philosophy of Modern Physics* **27** 47-79.